DOCUMENT RESUME

ED 187 331 IR 008 420

AUTHOR Falk, Joyce Duncan
TITLE The Bibliographic Databases in History.
PUB DATE Aug 79
NOTE 20p.; Paper presented at a Conference on Data Bases in the Humanities and the Social Sciences (Hanover, NH, August 23-24, 1979).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Classification; *Data Bases; *History; *Indexing; Information Processing; *Information Retrieval; Information Systems; Online Systems

ABSTRACT

This examination of abstracting-indexing services for the field of history focuses on Historical Abstracts (HA) and America--History and Life (AHL), and their relationship to the American Bibliographical Center's Subject Profile Index (ABC-SPIndex). The history, scope, selection criteria, and classification arrangements of the two databases are described, as is the indexing system of ABC-SPIndex. Sections also explore the processing of entries, the kinds of personnel involved in the bibliographical process, user input, and the cost effective aspects of HA and AHL. The concluding section provides a summary of major points. (FM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED187331

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE BIBLIOGRAPHIC DATABASES IN HISTORY

by

Joyce Duncan Falk

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Joyce Duncan Falk

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR008420

AUTHOR ABSTRACT

Author: Joyce Duncan FALK

Title: The Bibliographic Databases in History

Abstract: Historical Abstracts (HA) and America: History and Life (AHL), the abstracting-indexing services for the field of history, are available in printed and online versions (Lockheed's DIALOG). Recently expanded coverage includes books, articles, and dissertations. Bibliographic entries are classified and indexed according to the American Bibliographical Center's Subject Profile Index (ABC-SPIndex), which combines the skills of subject-specialist indexers and computer processing. ABC-SPIndex uses natural language keywords and a list of preferred subject headings in a rotated-string index. Each index entry is a string of several subject, biographic, and geographic terms followed by a date. In the printed index, each entry is rotated so that the entire string appears at each of the individual terms and appears as many times as there are terms in the string. HA and AHL depend upon an international group of volunteer and free-lance abstracters from the history and library professions and on a highly-trained staff of editor-indexers. Future refinements and expansions of both the printed and online versions will depend upon the desires of subscribers and financial resources available.

Paper presented at:

Conference on Data Bases in the Humanities and the Social Sciences, Dartmouth College, August 23-24, 1979.

"The Bibliographic Databases in History"

Joyce Duncan Falk, Director
American Bibliographical Center
(ABC-Clio, Inc.)

INTRODUCTION

Historical Abstracts and America: History and Life are the abstracting and indexing services for the field of history, available in both printed and machine-readable formats. These bibliographic databases are compiled and published by the American Bibliographical Center (ABC), a department of ABC-Clio, Inc., Santa Barbara, California.[1]

Bibliographic entries in Historical Abstracts and America: History and Life are of several types. 1) Articles from serials and collections are represented by full bibliographic data plus an abstract or annotation; occasionally when the title is fully descriptive, no annotation is included. 2) Book and dissertation entries include full publication data but no annotation; AHL includes book review data. 3) Approximately 300 journals of highly specialized or local subject matter are cited by journal title and are represented by a brief description of the subject content of the journal, called a journal profile. The language of the bibliographies is English but non-English original article titles and journal titles are included.

Historical Abstracts (HA), founded in 1954, is divided into two parts with four issues each per year: Part A, Modern History Abstracts (1450-1914) and Part B, Twentieth Century Abstracts (1914-present): each part consists of three abstracts issues and one cumulative index per year. The total number of entries in HA through volume 25 (1979) is more than 133,000.

America: History and Life (AHL) appears in four parts: Part A, Article Abstracts and Citations, three issues annually; Part B, Index to Book Reviews, two issues per year; Part C, American History Bibliography (Books, Articles, and Dissertations), annual; and Part D, Annual Index, the annual cumulative subject and author index to Parts A, B, and C, plus reviewer index, list of periodicals, and list of abstracters. Since the expansion with volume 11 in 1974, AHL has cited an average of 8100 articles, 2000 books, and 1500 dissertations per year. The total number of bibliographic entries through volume 16 (1979) will be 110,000. (This figure does not include second and later reviews of a book; book titles are counted only once for purposes of total number of bibliographic entries.)

Both HA and AHL are also available on Lockheed's DIALOG online information retrieval system. All volumes and parts of AHL are available, 1964 to the present, approximately 98,000 citations; the HA online file begins with volume 19 (1973) and has approximately 64,000 citations.

SCOPE

HA contains abstracts and shorter annotations of the periodical literature on world history of the period 1450 to the present, except on the U.S. and Canada. AHL covers the periodical literature on the history and culture of the U.S. and Canada, prehistory to the present (that is, without any chronological limitations on the subject matter), and since 1974 has also indexed books, book reviews, and dissertations. Beginning with volume 31 (1980), HA will add the indexing of books and dissertations to its coverage of articles. HA and AHL concentrate on literature in the field of history, but they also include material from the related social sciences and humanitites, so that in addition to being

the basic bibliographies for history, they supply supplementary support for the social sciences (such as anthropology, political science, sociology, and education), the humanities, area studies, and other interdisciplinary studies (especially American Studies, popular culture, and intellectual and cultural history).

To provide citations within this rather broadly defined subject scope of HA and AHL, editors regularly survey some 2000 serial publications, occasional books on collections of articles, and Dissertation Abstracts International (DAI).

Books listed in AHL are selected from the book reviews in a select group of 130 journals representative of various types of history and all regions of the U.S. and Canada. In addition to listing books on U.S. and Canadian history in Part C of its bibliographical system, AHL indexes the reviews of those books from 130 journals in a separately published Part B, Index to Book Reviews. Beginning in 1980, HA will list new books selected from the reviews in 10-15 English-language journals having the fastest and most comprehensive book review service; but HA does not plan to index the review data.

The same 2000 journals, occasional collections, and DAI are surveyed for both HA and AHL. A particular article, book, or dissertation entry is coded for listing in HA, AHL, or both, depending on its subject matter. The HA-AHL system aims to cover all the significant history journals of the world, the major political science, sociology, economics, and general social science journals of most countries, the history of disciplines such as the history of science and the history of education, and journals in the humanities that include historical material such as literary history and the history of philosophy; leading journals of

opinion and international affairs are also included in the list of periodicals surveyed; virtually all U.S. history publications beyond mere newsletters are included, but no attempt is made to cover all the local history and county or provincial history journals of other countries.

In addition to being limited to a specific list of periodicals and other sources and to the subject scope of history, broadly defined, articles selected must be at least three pages long (except obituaries of historians). Book reviews are not listed but review articles are included; HA and AHL exclude editorials and opinion or propaganda pieces without at least 20 percent historical content. Evaluative judgments of the books, articles, and dissertations are not a basis for selection of the items.

The subject scope and specific selection criteria have evolved over the years. For example, in 1975 a substantial change in the scope of HA was made when editors realized that HA could not cover all the historical material within scope and at the same time continue to include the great quantity of current affairs and opinion that was being listed. Although the policy of including history "to the present" makes selection of material on recent events difficult, editors have decreased the number of items listed in HA that merely report or comment on contemporary events and issues. At the same time, guidelines are more strict than in 1974 and earlier for inclusion of social science and other peripherally related material, especially sociology, education, and library science. On the other hand, HA editors no longer try to identify and exclude local history articles; HA excludes local history journals but includes articles on local history from the journals surveyed.

Traditional political, diplomatic, and military history has been balanced by the addition of journals and by selection of articles on social, cultural, and intellectual history. Editors are encouraged to be sensitive to developments in the field such as the serious scholarly interest in ethnic studies, family history, and popular culture.

The scope and selection criteria are determined by the editorial staff of ABC, the majority of whom are trained in history or closely related subjects. They take into account established philosophy of HA and AHL and the history or traditions of these two bibliographic enterprises, the present trends in the profession, and the practical necessities of bibliographical control and financial viability. The director of ABC approves the general guidelines and specific criteria and serves as the final authority on questions of interpretation of selection policies as applied to journals to be covered and individual articles, books, and dissertations to be selected. Actual selection is done by three groups of persons: more than half of the material is selected by in-house editors--the same editors who work on the publication. Many of the non-English-language journals are screened by local multilingual scholars. In a few instances, particularly with journals in very unusual languages or in cases where we do not have the journal issue and have to use a library copy, we ask the abstracter to select the relevant articles.

Thus, the scope of HA and AHL is determined by experienced editors and implemented, for the most part, by them. Selection of individual items is made from a finite list of serials and other sources. Coverage is comprehensive within those sources, limited by subject scope and certain objective criteria, and not based on qualitative assessments

of the items. Interpretation of the subject scope and even of the more specific, objective criteria is, however, a matter of judgment.

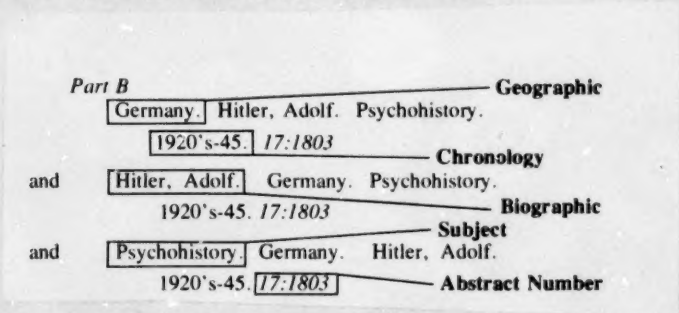
CLASSIFICATION AND INDEXING

One means of access to the bibliographic entries is through the classified arrangement. The entries are classified according to geographical-chronological-topical schemes developed especially for HA and AHL by in-house editors with the advice of teachers and scholars in the field of history. The printed versions of the databases are arranged according to the respective classification schemes (shown in the table of contents) and the online files are searchable by classification codes and headings. The classification systems are independent of the indexing. They are reviewed at five-year intervals and although the temptation is always to improve the schemes in several respects, because the classification codes are used in online searching, we limit changes primarily to extensions or further subdivisions of the existing divisions.

The more specific means of access to the bibliographic entries is the subject indexing. In 1974-75 the editorial staff of ABC developed the Subject Profile Index (ABC-SPIndex) for the history databases.[2] (An earlier, simpler version was being used in Environmental Periodicals Bibliography, published by the International Academy of Santa Barbara but the principles, procedures, and processing of ABC-SPIndex differ markedly from that version.)

Entries for articles are indexed on the basis of the title and abstract; occasionally, if the title and annotation are not sufficiently descriptive or clear, editors consult the article itself. Book reviews are the basis for indexing books, and DAI abstracts for indexing dissertation entries.

Each index entry is a string of terms: an average of 3.7 subject, biographic, and geographic descriptors. In addition, each index string includes a chronologic descriptor--a date or set of dates--and, finally, the reference number of the bibliographic entry. The index entry thus provides a subject profile of the original document by listing together all its index terms including the historical period (date). The entries in the online database carry these terms in the descriptor field. In the printed format, the string of terms is rotated by computer processing so that the entire string--the subject profile--of the book, article, dissertation, or journal appears under each of the individual terms in the string and appears as many times as there are terms.



The indexing vocabulary is a hybrid of natural language keywords and preferred subject headings. By natural language I mean the language appearing in the title or abstract, as opposed to assigned subject headings. Two captivating, but not entirely compatible, notions of using the computer to achieve economies in human labor strongly influenced the development of ABC-SPIndex. The ability to tag words in the title and abstract for transfer to the index without additional keyboarding promoted the concept of a natural language indexing system. Another device, the use of codes or abbreviations for frequently used index terms to be converted by the computer to full-text terms (thus saving writing out

the entire term and keyboarding it), led to the establishment of subject headings.

A list of some 1900 headings with codes, called "cues," was drawn up based on Sears List of Subject Headings and the existing HA-AHL lists of geographical terms and major historical events. It included "see also" references and a separate list of "see" references. Use of this list together with the editors' dissatisfaction with the scattering and inconsistency evident in the first issue of AHL using ABC-SPIndex led to more control and less tagging of free language.

ABC-SPIndex terms are selected by editor-indexers by three means: 1) Keywords or phrases from titles and texts of abstracts or annotations are tagged and then automatically set aside by computer processing to become part of the index string. These words and phrases may be free, natural language terms or they may be words that are identical to the subject headings. This is the most economical of the three methods of marking index terms. 2) Independent terms may be written onto the input form by the editor and keyboarded in full text. This is the least used type of index term, but it is sometimes necessary in order to get the most appropriate term or the correct word order, punctuation, or typeface. 3) Cues (abbreviations or codes) are selected from the preferred subject heading list, written onto the input form, and keyboarded; the cues are later converted by the computer to full text and are joined with terms from other sources to form the index string.

An abstract or other bibliographic entry may be indexed using any or all of the three methods of denoting index terms. About 27-29 percent of the index terms are from cues; or approximately one index term per bibliographic entry is a cue. Tagged terms make up about 50-53 per-

cent of the terms and independently added, full-text words or phrases account for about 20 percent. These percentages, however, underestimate the weight of the preferred subject headings because tagged terms can be identical to those headings. In fact, the wording of an abstract is sometimes changed to conform to the established subject heading to facilitate tagging. Eventually all three types of terms are merged to form the individual index strings for each bibliographic entry, and neither the printed index nor the online file makes any distinction among the three.

ABC-SPIndex is basically a natural language, uncontrolled vocabulary indexing system, but it does have precoordinated terms, some control exercised through authority files, and syndetic structure. It is open-ended and dynamic: a new index term can be added at any time, provided it does not violate existing authorities. Since its compilation in 1974, the list of subject headings and cues has been slightly revised twice.

ABC-SPIndex does not indicate the relationship between any two terms in the index string, but some relationships are shown through precoordinated subject headings and qualification of terms by adding modifiers in parentheses or by inverting words in a phrase.

Index terms may be one word or several; they are mostly nouns and proper names, but also gerunds and modified nouns or precoordinated terms like "class struggle," history teaching," "economic development," and "antifascist movement." Most multiword terms are in direct order, but some are inverted ("uniforms, military," "education, history of," and "Europe, central"). The early history of ABC-SPIndex is evident in the continued presence of a few compound headings from Sears ("missions

and missionaries," "lumber and lumbering"). These terms, awkward in ABC-SPIndex and contrary to its basic principle of highly specific indexing, are gradually being replaced by single terms.

Index terms may be modified as the need arises. Smaller geographical units such as cities, towns, counties, and provinces always serve as modifiers of the larger entity to which they belong. In AHL, the states are used as leading terms, smaller geographical subdivisions as modifiers; in HA, countries are the leading terms and states, provinces, regions, towns, etc., are modifiers: "France (Normandy)" or Spain (Madrid)." Modifiers are used to clarify ambiguous words and to supply additional information: "Echo du Nord (newspaper)," "International Relations (discipline)," and "Bismarck (vessel)." Both the subject headings and independent full-text terms can be modified by the addition of words in parentheses. For the index, tagged words from the title and abstract can be inverted by inserting a particular code.

Some restraint on this freedom and flexibility is exercised through several authority devices and a few general rules. The basic rule is to index at the level of specificity of the original document or abstract and to use the most specific terms applicable.

The first authority device is the list of prohibited terms, the see-reference list ("Bourgeoisie, see Middle Classes"; "Underdeveloped Nations, see Developing Nations"). The second authority is the list of preferred subject headings. This list provides the cue to be used instead of the full-text words and it shows the see-also reference. When several synonymous or closely related terms are possible as index terms, the term in the subject heading list is used. A third authority is the previous annual index of the particular publication--although

this authority is not foolproof or absolute. The fourth authority is a file of authority cards on which are registered decisions about the correct entry form to use for people, events, and subjects that have various possible names. Editors add to this file continually; two sets of the cards are available in the editorial offices. Future plans are to combine and print the subject heading list, augmented with scope notes, and the authority file in a single book, to be available on each editor's desk.

The terms in the index string sort alphabetically following the leading terms. In the example above (page 7), the first string is re-sorted at each appearance so that the second and following terms are in alphabetical order within the string. This sorting has the advantage of pulling many similar strings together, but it can be misleading: not all identical terms sort together at the second position in the index string. For example, note the position of "Communism" and "Political Theory" in index strings beginning with "Lenin":

Lenin, V. I. Antipolitik. Communism. Political Theory. 19c-20c. *1262*
—. Class struggle. Ideology. Leftism. 1919-20. *224*
—. Communism. Patriotism. Political Theory. 20c. *231*
—. Political Theory. Self-determination. 1900-24. *2252*

The cumulative annual index differs from the issue indexes in that it summarizes the leading terms (i.e., uses a dash in place of repeated identical leading terms). The annual index also includes the cross references, which are not present in the individual issue indexes. See-also references are generated by the system from ones stored in the subject heading file and verified by the editors. Because the index terms are derived from the text and independently added terms and not exclusively from the file of subject headings, and because no diction-

aries are used to convert various forms of a word to a single form or to treat various forms as if they were the same word, the procedure for producing cross references is not completely automated. Editors manually compare a leading term list (a list of all the words used as index terms) for the particular volume to the master subject heading file and mark the terms to be included as cross references in the annual index. The cross references are then generated from the "corrected" master subject heading file. See references are inserted from a separate file and are edited as appropriate for the particular volume. At the present time the system does not use stop lists or other devices that prevent the accidental use of a prohibited word from the see-reference list as an index term. Editors must catch such errors during the editing of the annual index and verification of cross-references.

In addition to the Subject Profile Index, each issue of HA and AHL and every cumulative annual index includes an author index. AHL Part B, Index to Book Reviews contains a reviewer index and, beginning with Volume 16 (1979), a title index. Five-year cumulative indexes are also published. HA volumes 16-20 and AHL Volumes 6-10 were completely reindexed using ABC-SPIndex, so the five-year indexes to those volumes offer the same kind of subject access as current volumes.

PROCESSING

The ability to use a computer to manipulate large amounts of data made the idea of a rotated string index feasible. This manipulation, especially sorting, together with the tagging and cueing procedures mentioned above and with photocomposed typesetting of camera-ready pages are the basic features of the HA-AHL computerized processing system.

Edited, classified, and indexed bibliographic entries are submitted

to the ABC-Clio Data Processing Services Department (DPS) for keyboarding and transfer to magnetic tape files. (DPS currently uses Compugraphic MDT 400 terminals.) Keyboarding includes insertion of numerous codes that, after further processing, result in the creation of subject and author indexes and the production of pages for printing. Both author and subject index terms are typed at the same time as the bibliographic entry itself, i.e., author names and those words in the title or annotation that are tagged for index terms are typed only once, with the appropriate code. Codes for the various fields of data (classification category, author, title, journal, abstract, abstracter, cues, independently added index terms, etc.) and codes to effect diacritical marks and changes of type face are inserted by the typist. The magnetic tapes are processed on an Interdata 7/16 minicomputer or an IBM 370. Listings of bibliographic entries and indexes are produced on a line printer and proofread, copyread, and further checked. Corrections are made to the computer files (tapes or disks) via CRT terminals. The precise configuration of equipment and procedures and the use of out-of-house contractors changes from time to time.

After the computer files are corrected, the tapes drive an RCA Videocomp to produce photocomposed pages ready to be printed via lithographing after editors do final checking of pages and arrange for paste-up of corrections and front matter. The same tapes used for typesetting, but usually subjected to another correction pass, are sent to Lockheed for loading onto the DIALOG system.

PERSONNEL AND USER INPUT

The network of persons who contribute to HA and AHL stretches around the world and includes a variety of relationships. The ABC

library staff (one professional librarian and four paraprofessionals) recruit abstracters, obtain journals and other source documents, send journals and instructions to abstracters, and maintain the records on journals and bibliographic entries (document control).

Originally based almost entirely on volunteer abstracters, the HA-AHL system still depends upon nearly 400 historians and librarians to prepare abstracts of articles assigned by the library staff. To achieve greater currency in publishing the bibliographic entries and to catch up on a backlog of approximately 30,000 articles, part-time free-lance abstracters and bibliographic assistants, some working directly for the Santa Barbara office and about 80 working for the HA office at the European Bibliographical Centre in Oxford, England, now supplement volunteer efforts. These assistants are primarily scholars with advanced degrees in history or related areas; a few local graduate students and other persons prepare some of the bibliographic entries with shorter annotations. Area services at several universities supply the abstracts for certain assigned journals. Presently there are area services at the University of Leeds, University of Salzburg, University of Washington, and the University of Tokyo with the assistance of the Japanese Historical Society.

Book and dissertation entries and journal profiles are prepared by an in-house editor or by a former staff member.

The 14-member in-house editorial staff of HA and AHL including four editors in Oxford, England, are responsible for editing, classifying, and indexing the bibliographic entries. They work first on the manuscript abstract or other original entry, later on listings of the entries and indexes produced by the computer via a line printer, and

finally, on the camera-ready pages. Editors are mostly subject specialists, that is, they have degrees in history or closely related subjects. Several persons with B.A.'s in English have begun as editorial assistants and some have moved to assistant editor positions. All have some facility in a foreign language, mostly the Romance languages, but some German and Slavic as well. Very occasionally a person with unusual language skills, but without training specifically in history, works as a free-lance abstracter or bibliographic assistant. The director of the American Bibliographical Center, who is executive editor of HA and AHL, has a background in both history teaching and librarianship.

In addition to the subject expertise of the editorial staff, contact with the history and library professions is maintained through the abstracters and through an international advisory board of distinguished scholars for both HA and AHL. Further user feedback and advice is sought through personal conversations with librarians and historians at national and regional professional meetings and by questionnaires to subscribers, advisory board members, and abstracters. During 1979 an HA Users Board of eighteen librarians and other interested persons was convened to consider a number of questions about the expansion of HA.

COSTS

The printed versions of HA and AHL are priced on a service rate basis; the subscription rate varies according to the library's budget for books, periodicals, and binding. Individuals who are members of the Organization of American Historians can subscribe to AHL at a special reduced rate based on print overrun costs. The HA and AHL databases can be searched on DIALOG for $65 per hour and $0.15 per printed citation

(with abstract or book review data where it exists), plus communication costs. An average search in HA or AHL costs $15-$20.

Because neither the initial development nor continuing production of HA and AHL are subsidized by tax monies, foundation grants, or larger corporation tax-write-off investments, improvements to the databases, whether in printed or online format, must be financed by subscription income. Future refinement of the indexing system, modifications for online use, currency and extent of coverage, and more extensive, sophisticated use of the computer for quality control depends on the desires and financial ability of the HA and AHL subscribers.

CONCLUSION

To meet the demands for a comprehensive and current bibliographical system in history, adequately indexed, HA and AHL now relies less on volunteer abstracters and more on free-lance and in-house staff. The American Bibliographical Center is pursuing an extensive effort to improve the bibliographic control of the literature of history by catching up on the backlog of bibliographic entries not yet published and by expanding the coverage of HA and AHL.

HA and AHL are labor intensive publications. Computerization of some of the processes has added expense and processing time without decreasing labor costs. ABC-SPIndex combines human indexing skill with computer processing to produce a highly specific, informative index for HA and AHL, but not a less expensive or faster one.

Research and development are pursued by the editorial and production staffs while they produce the publications. ABC-SPIndex was developed during use, without prior study, design, and testing. Once the publisher decided on the principle of a rotated string of index terms derived

from tagged words supplemented by cues for subject headings, the editors first of AHL and then of HA, during the editing of the first issues to use ABC-SPIndex, drew up rules for implementation, subject headings, cross references, and authority cards.

ABC-SPIndex is a hybrid of keyword, natural language, and assigned subject-heading indexing. The indexing vocabulary is basically uncontrolled, but some control is exercised through the use of authority cards and preferred subject headings.

NOTES

1. Joyce Duncan Falk, "Controlled and Free Vocabulary Indexing of the ABC-Clio Databases in History," in Proceedings of the Conference of Data Bases in the Humanities and the Social Sciences, Dartmouth, August, 1979, ed. J. Raben and G. Marks (Amsterdam: North Holland, forthcoming July 1980). The present paper is an expanded version of that article; the introductory section and the section on classification and indexing below are copyrighted in that earlier article.
2. Joyce Duncan Falk, "Introduction," America: History and Life Five Year Index Vols. 6-10 (Santa Barbara: ABC-Clio, 1977) and "Introduction," Historical Abstracts Five Year Index Vols. 16-20 (Santa Barbara: ABC-Clio, 1978).